Computational models investigating the relative benefits of ensembles of words on early word recognition for learners of different levels of reading skill

It is taken as a truism that the learning environment matters for early reading development. This fact bears out in a number of indirect ways, including treatment effects associated with experiments on different forms of reading instruction (Foorman et al., 1998) and differences in processing difficulty of words as measured in psycholinguistic megastudies (Balota et al., 2007), among others. This assumption also pervades code-based instructional programs that are designed to teach children how to develop word recognition skills. Words in these programs are carefully curated for various properties of print and speech, with special consideration of children with slow developing decoding skills.

However, identifying the learning benefits of specific words (and ensembles of them) for children’s word recognition skills is difficult. This is due to several factors, including variability in children’s language knowledge (including for print vocabulary), limited instructional time, and general challenges associated with behavioral experimentation with young children in the early elementary years. This study sought to investigate the issue using computational models within the triangle framework (Seidenberg & McClelland, 1989) in order to understand the extent to which specific sets of words can differentially effect outcomes for learners of different skill levels.

Using connectionist models that learn to name printed words, we constructed random training environments for a large number of simulated learners (*N* = 50,000), each trained to produce named responses for all words in a learning environment of 300 words. In order to investigate the interaction of print environment and skill level, models varied in their capacity to learn print-speech mappings, which was accomplished by manipulating the number of interlevel units contained in the model from 20 to 100 units. Results indicate dramatic variation among learning environments constructed with different ensembles of words, and these differences are pronounced across levels of learner skill. For example, impaired models are associated with an average of 1.59 lower accuracy (in SD units) than their unimpaired counterparts across all 300 training words. Even more dramatic are the relative outcomes of the most and least successful learners across the two groups, where differences exceed 5 standard deviation units. These results have implications for the construction of print learning environments (including but not limited to instruction) for early developing readers across different levels of skill and help frame accounts of differential benefits of words and their structural properties on learning.